

~~PROPOSED~~ ORDER RE: CONVERSATION DATA SAMPLING

It is now ORDERED, that the following procedures shall apply to the sampling of the consumer output log data set forth in the Court's May 29, 2025 Order:

1. As soon as practicable following entry of this Order, OpenAI shall generate data samples using a random sampling methodology as follows:
 - a. Each sample shall consist of the contents of individual rows in the table of consumer output log data (the "Preserved Table") that OpenAI segregated and preserved in response to the Court's May 13, 2025 Order, MDL ECF No. 33 (the "Preservation Order").
 - b. OpenAI shall use best efforts to include in each sample data rows (i) corresponding to outputs generated between April 14, 2025 and May 14, 2025; (ii) corresponding to interactions with ChatGPT Free, Plus, or Pro using the models GPT-3.5, GPT-3.5 Turbo, GPT-4, GPT-4 Turbo, GPT-4o, and GPT-4o Mini; (iii) corresponding to user requests originating from outside the European Economic Area, Switzerland, or the United Kingdom; and (iv) where the values of the conversation classifier fields included in the Preserved Tables are not "null." The filtered data in the Preserved Table shall hereinafter be referred to as "Conversation Data."
 - c. OpenAI shall use best efforts to draw independent and identically distributed samples from the sample populations described below:
 - i. The sample population for the first sample shall be all Conversation Data that (i) was not generated through the "Temporary Chat" feature and (ii) is not subject to a user-initiated deletion. *See* MDL ECF 66 ¶ 3.
 - ii. The sample population for the second sample shall consist of Conversation Data that was generated through the "Temporary Chat" feature. *See* MDL ECF 66 ¶ 3.
 - iii. The sample population for the third sample shall consist of Conversation Data that was subject to a user-initiated deletion. *See* MDL ECF 66 ¶ 3.
 - iv. The sample population for the fourth sample shall be all Conversation Data.
 - d. Solely for purposes of complying with this Order, the size of each data sample shall be five (5) million rows of Conversation Data.

2. Within seven (7) days of generating the samples, OpenAI shall run searches on each sample to generate the reports described in Paragraph 3.
3. After running the searches in the foregoing paragraph, OpenAI will provide News Plaintiffs¹ with, for each of the four samples specified in Paragraph 1(c):
 - a. Counts of the number of rows that contain a match in the prompt or output field for each of the search terms listed at MDL ECF No. 404-2 at pp. 5-8, and
 - b. Counts of the number of rows that contain a match in the `classification_capability_requested` field for the classification categories listed below (with counts provided separately for each of the classification categories):
 - i. `info_about_public_people`
 - ii. `info_about_current_events_and_news`
 - iii. `purchasable_products`
 - iv. `cooking_and_recipes`
4. The data described in Paragraph 3 above, and any analyses derived from the data, shall be designated as CONFIDENTIAL pursuant to the operative Protective Order in this action.
5. Absent further order from the Court, all data and analyses derived from the samples, including hit count tables, may only be used for the sole purpose of preparing the joint letter described in Paragraph 6 below, and any related analyses associated with this sampling exercise.
6. Seven (7) days after OpenAI provides the hit count tables for the samples described in Paragraph 4 above, OpenAI and News Plaintiffs shall submit a joint letter of no more than three (3) pages (no more than one-and-a-half (1.5) pages per side, not including any exhibits) that attaches the hit count tables and presents each side's position regarding the sampling exercise.
7. Seven (7) days after the parties submit the joint letter described in Paragraph 6 above, or at the Court's earliest convenience, the Court shall convene a settlement conference to discuss the results of the sampling exercise and the impact on the Preservation Order.

¹ "News Plaintiffs" shall collectively refer to Plaintiffs The New York Times Company; Daily News, LP; Chicago Tribune Company, LLC; Orlando Sentinel Communications Company, LLC; Sun-Sentinel Company, LLC; San Jose Mercury-News, LLC; DP Media Network, LLC; ORB Publishing, LLC; Northwest Publications, LLC; and the Center for Investigative Reporting.

SO ORDERED.



Ona T. Wang
U.S.M.J. 8/19/2025